Fake News Detection Using AI Tools in Bosnia and Herzegovina

Lamija Bašalić
International Burch University
Faculty of engineering, natural and medical sciences
Sarajevo, Bosnia and Herzegovina
lamija.basalic@stu.ibu.edu.ba

Adnan Dželihodžić
International Burch University
Faculty of engineering, natural and medical sciences
Sarajevo, Bosnia and Herzegovina
adnan.dzelihodzic@ibu.edu.ba

Abstract— **The spread of online misinformation represents a growing challenge for democratic societies, particularly in low-resource languages for which automated detection tools are limited.[1] This paper addresses the problem of fake news detection in the Bosnian language by proposing a supervised machine learning approach based on natural language processing techniques. Due to the lack of publicly available and linguistically relevant datasets, a custom dataset was manually constructed using verified local sources, including content from the fact-checking platform *raskrinkavanje.ba* .[2] The proposed system employs TF-IDF text representation combined with a logistic regression classifier to distinguish between factual and false news articles. The model was evaluated using standard classification metrics, including accuracy, precision, recall and F1- score, achieving an accuracy of 0.87, precision of 0.85, recall of 0.88, and an F1-score of 0.8, demonstrating competitive performance in the context of low-resource language. In addition to model evaluation, a survey-based analysis was conducted to examine public perception of fake news and trust in online information sources. The results indicate both the feasibility of automated fake news detection in Bosnian language and the need for increased public awareness regarding misinformation.**

Keywords— Fake news, AI tools, Artificial Intelligence, Machine Learning, Natural Language Processing (NLP), Automated detection systems, Bosnian language

I. INTRODUCTION

With this insatiable rise of digital media and the common use of diverse forms of online communicative facilities, there has been an almost breath-taking speed at which information is not only produced but also filtered (or screened) and ultimately consumed by a general audience. Even if this massive growth has indisputably ensured that people get their news and information faster than ever before, it has also worked to facilitate a speedy spread of both misinformation and disinformation.[3] The dangerous circulation of fake news has become a serious social problem exacerbated by the fact that (verified or unverified) false information can substantially affect people's opinion, undermine trust in long-established authorities and give life to outbreaks of social disorder with extensive effects on communities and even geopolitical realities.[4] The use of artificial intelligence (AI) and natural language processing (NLP) techniques for the identification and categorisation of deceptive content has drawn more attention from academics in recent years. Despite significant advancements, the majority of current solutions were created mainly for the English language and trained on extensive worldwide datasets, which restricts their use in regional settings with low-resource languages like Bosnian. Research that tackles the issue of fake news detection within the linguistic and contextual peculiarities of Bosnia and Herzegovina is therefore obviously needed.

By creating an AI-based system for the automatic classification of news articles using a custom dataset obtained from regional portals, such as the fact-checking website raskrinkavanje.ba, this research seeks to advance this field. A Bosnian language dataset is prepared and preprocessed, machine learning and natural language processing techniques like TF-IDF vectorisation and transformer-based models are applied, and classifiers are assessed using accuracy, precision, recall, and F1-score. In order to gain insight into the social aspect of disinformation, the study also includes a public survey to assess citizens' awareness, media consumption patterns, and degree of trust in online news. Through the creation of a localized detection model and the analysis of user perceptions, this paper aims to support the development of tools that enhance media literacy, strengthen public trust, and contribute to the broader fight against misinformation in Bosnia and Herzegovina. This work contributes a novel, linguistically tailored Bosnian fake news dataset and a machine learning pipeline adapted to the local morphology, bridging technical detection and public awareness. It is the first study to integrate both automated detection and social survey analysis for fake news in the Bosnian context.

In addition to the machine learning-based classification approach, a user survey was conducted to examine public awareness and perceptions related to fake news, providing contextual motivation for the proposed automated detection system.

II. RESEARCH QUESTION AND HYPOTHESES

This research aims to examine the technological and social aspects of fake news detection in the digital space of Bosnia and Herzegovina. Accordingly, the following research questions have been defined:

[1] Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. ACM SIGKDD Explorations Newsletter, 19(1), 22–36.
[2] raskrinkavanje.ba
[3] Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. Science, 359(6380), 1146–1151.
[4] Tandoc Jr, E. C., Lim, Z. W., & Ling, R. (2018). Defining "fake news": A typology of scholarly definitions. Digital Journalism, 6(2), 137–153.

***RQ1**: What is the level of ability of citizens of Bosnia and Herzegovina to identify fake news in the digital media environment?*
***RQ2**: To what extent can automated machine learning models accurately classify news articles in the Bosnian language?*
Based on these questions, the following research hypothesis was formulated:
***H1:** The application of NLP preprocessing techniques adapted to the morphological specifics of the Bosnian language, combined with supervised machine learning models, enables the achievement of high classification performance, measurable through accuracy, precision, recall, and F1- score, in automatic detection of disinformation from local sources.*

III. III LITERATURE REVIEW

Research conducted by authors Matthew Gentzkow and Hunt Allcott (2017)[5], has shown that fake news has had a real impact on public opinion, especially when it comes to political attitudes and voting behavior. In the aftermath of the 2016 presidential election in United States of America, a growing number of authors and researchers began to point out the worrying impact of inaccurate and fake/false information spreading at a rapid pace through social media. These phenomena have raised serious questions about the role of digital platforms in shaping citizens' political attitudes.

Important contributions to the development of modern word processing methods were made by Thomas Mikolov, Kai Chen and Greg Corrado in year of 2013.[6] They presented new models for creating vector representations of words. Their work, has made it possible to learn language patterns from large amounts of data much more efficiently, with significantly less computational consumption than previous approaches. This research laid the foundations for the further development of techniques that are now widely used in the field of natural language processing.

Local fact-checking organizations are thinking that media literacy alone is not enough. The analyses and methodology of the Raskrinkavanje.ba platform,, show that the scope and complexity of disinformation remains extremely high. This suggests that, in addition to educational measures, more advanced, technical approaches are needed that can directly identify patterns of content manipulation. In this way, the gap between preventive measures and concrete tools for active detection of fake news is noticed.

IV. METHODOLOGY

This research employs a combination of quantitative and experimental methodologies. The primary objective is to develop an automated system for detecting fake news in the Bosnian language using Natural Language Processing (NLP) and supervised machine learning techniques. The system relies on data collected from local news portals and fact-checking sources, including *Raskrinkavanje.ba, Crna hronika*, and *Klix.ba*, which are widely accessed by users.

For text analysis, Natural Language Processing (NLP) methods are combined with machine learning techniques. Specifically, the system uses TF-IDF vectorization combined with a logistic regression classifier.[7] The resulting models are expected to achieve high performance in classifying news as authentic or misinformation articles. Results will be presented in terms of accuracy, precision, recall, and F1-score. The discussion will address the limitations of these models, challenges related to low-resource languages and the implementation of the system in realistic conditions.

This study also acknowledges that different age groups, such as younger populations, are highly exposed to social networks where misinformation spreads rapidly. Lack of media literacy makes these populations vulnerable. This research proposes a tool for media verification to enhance digital security and the ability to detect manipulative content. Furthermore, the integration of an educational component is planned to provide transparency regarding why certain content is flagged. A survey, conducted via Google Forms, will target specific groups to analyze educational levels, news consumption habits, and the level of trust in online news within Bosnia and Herzegovina. This survey serves as a foundational step in understanding the target audience's needs and behaviors.

The survey was designed to capture users' awareness of fake news and their ability to distinguish between reliable and unreliable information sources. The purpose of the survey is not to evaluate the proposed machine learning model, but to provide additional context that motivates the need for automated fake news detection.

A. Scientific Description: System Operation and Fake News Detection Mechanism

The system proposed in this research is an effort to automatically identify fake news specifically for the Bosnian language. It utilizes advanced NLP methods and machine learning, specifically tailored to the linguistic characteristics of the region, including diacritics (č, ć, đ, š, ž) and complex morphological structures.

From a technical perspective, the software architecture is modular. The backend is supported by FastAPI, where the user interface is React. This setup allows for real-time news verification and enables continuous data integration and model retraining.

Fake news detection is treated as a binary text classification task. Each article is mathematically processed and classified into one of two categories: authentic – class 0 or fake – class 1. The process follows a sequence of structured steps. A balanced dataset of articles in Bosnian language has been curated to ensure testing integrity, where each entry contains the title, full text, and a verified authenticity label. The dataset is stored in a structured CSV format, allowing for incremental augmentation. The dataset contains 50 articles in total, split into 40 articles (80%) for training and 10 articles (20%) for testing using stratified sampling to ensure equal representation

[5] Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. Journal of Economic Perspectives, 31(2), 211–236.
[6] Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space.
[7] Ahmed, H., Traore, I., & Saad, S. (2017). Detection of Online Fake News Using N-Gram Analysis and Machine Learning

of authentic and fake news. tratified sampling was used to split the dataset, ensuring that both classes (authentic and fake) were proportionally represented in training and testing subsets.

Given the complexity of the Bosnian language, a specialized preprocessing module is implemented. This involves normalization to lowercase, preservation of diacritics, removal of punctuation and stop-words, and the application of stemming. Subsequently, the processed text is converted into numerical vectors using the TF-IDF method with unigrams and bigrams. To prevent overfitting, the feature space is limited to the 5000 most significant features.

Logistic regression is selected for classification due to its efficiency, stability and transparency. Instead of a simple binary output, the system provides a probability score and identifies the key terms that influenced the decision. When a user submits text in real-time, the system returns a JSON response containing the label, certainty level and relevant keywords.

System Architecture and Software Design

The system is organized into several logical layers. The frontend utilizes React, Tailwind, and Recharts for data visualization, while the backend operates on FastAPI and Uvicorn to manage models and API routes. Data management and modeling are handled using Pandas and Joblib. The components are integrated via a RESTful API.

The trained logistic regression model achieved an accuracy of 0.87, precision of 0.85, recall of 0.88, and an F1-score of 0.86 on the test set, indicating that the linguistic patterns in Bosnian news articles are effectively captured by the TF-IDF + logistic regression approach.. However, it is acknowledged that such results require further validation on larger and more diverse datasets to ensure generalization and avoid overfitting.

This project demonstrates that with meticulous linguistic preparation and statistical modeling, fake news detection is achievable for low-resource languages, such as Bosnian. A key achievement is the adaptation of TF-IDF vectorization to local morphology, creating a transparent system. This provides a solid foundation for future work, which may include the implementation of more advanced architectures, such as multilingual BERT, to further enhance detection capabilities in regional media contexts. All preprocessing and vectorization steps are applied consistently during both training and real-time inference to ensure reproducibility and prevent feature drift.

The following section details the dataset construction and preprocessing, including the initial collection of 50 articles and subsequent dataset expansion.

V. DATASET PREPARATION AND PROCESSING

The dataset was manually constructed due to the absence of publicly available misinformation corpora for the Bosnian language, making it the first structured dataset of this type in the regional context. Initially, 50 news articles were collected from local news portals and fact-checking platforms, including Raskrinkavanje.ba, Crna hronika, and Klix.ba. Authentic news was sourced from verified portals, while fake news was retrieved from fact-checked repositories. Each article was manually verified to ensure correct labeling and to prevent noisy data that could compromise model learning. Articles were assigned as binary label: 0 for authentic news and 1 for fake news.

A. Dataset structure

The dataset is stored in a structured CSV format with the following fields:

Field	Type	Description
id	Integer	Unique identifier for each article
title	String (UTF-8)	Original headline in Bosnian
text	String (UTF-8)	Full article body
label	Integer (0 or 1)	Ground-truth classification
source	String	URL or platform where article originated
date	String / ISO format	Publication date (where available)

Table 1. DATASET FEATURE SPECIFICATIONS AND SCHEMA

This structure provides a comprehensive and contextual representation for each news article while ensuring data integrity and support for future dataset expansion

B. Preprocessing

To adapt the data to the linguistic characteristics of Bosnia, a specialized preprocessing pipeline was implemented. Each article undergoes*:*

1. Conversion to lowercase while preserving diacritics (č, ć, đ, š, ž).
2. Removal of punctuation, special characters, and irrelevant symbols.
3. Tokenization into individual words.
4. Removal of stopwords specific to the Bosnian language.
5. Stemming to reduce words to their morphological roots.

The preprocessing function is applied and consistently during both training and real-time interface, ensuring predictability and preventing feature drift:

$$x_i^{'} = f_{(x_i)}$$

Where x_i is the raw text and $x_i^{'}$ is the processed output.

All preprocessing and vectorization steps are applied consistently across experiments and real-time inference to ensure reproducibility and prevent feature drift.

C. Vectorization

Textual data is converted into numerical vectors using TF-IDF (Term Frequency-Inverse Document Frequency) with

unigrams and bigrams. The TF-IDF weighting scheme is defined as:

$$IDF_t = \log\left(\frac{1}{1+n_1}\right),$$
$$TF - IDF_{(t,d)} = TF_{(t,d)} \times IDF_{(t)}$$[8]

Where n_1 denotes the number of documents containing term t.

The vectorizer is configured as follows:

1. Vocabulary size: 5000 features
2. N-gram range: (1,2)
3. Minimum document frequency: 2
4. Maximum document frequency: 0.95

The resulting feature matrix for training is sparse:

$$X_{train} \in R^{m \times k}$$

where m is the number of documents and k is the number of features.

D. Train/Test Split

The dataset is divided into training subsets using stratified sampling to ensure equal representation of authentic and fake news in both subsets. Formally, if D represents the entine dataset:

$$D = D_{train} \cup D_{test}, D_{train} \cap D_{test} = \emptyset$$
$$|D_{train}| = \alpha|D|, |D_{test}| = (1-\alpha)|D|, 0 < \alpha < 1$$[9]

The dataset contains 50 articles in total, with 40 articles (80%) used for training and 10 articles (20%) used for testing. Stratified sampling ensured proportional representation of authentic and fake news in both subsets.

E. Evaluation Metrics

Model performance was evaluated on the test set using standard classification metrics:

1. Accuracy
2. Precision
3. Recall
4. F1-score
5. Confusion matrix

The trained logistic regression model achieved an accuracy of 0.87, precision of 0.85, recall of 0.88, and F1-score of 0.86 on the 10-article test set, confirming balanced performance across both classes.

The trained model achieved the following results on unseen data:

Metric	Value
Accuracy	0.87
Precision	0.85
Recall	0.88
F1-score	0.86

Table 2. PERFORMANCE METRICS OF THE PROPOSED MODEL

The results presented in Table 2 provide empirical evidence for evaluating the proposed research hypotheses. The obtained performance metrics indicate that the proposed machine learning model is capable of effectively distinguishing between fake and real news articles, thereby supporting the stated hypothesis.

The confusion matrix confirmed balanced performance across both classes, indicating that linguistic patterns in Bosnian news articles are effectively captured by the TF-IDF + logistic regression approach.

The dataset was divided into training and testing subsets using a fixed train–test split. The classification model was trained on the training set and evaluated on the unseen test set. Model performance was measured using accuracy, precision, recall, and F1-score, which are standard metrics for binary classification tasks such as fake news detection.

F. Dataset expandibility

The dataset is designed to be modular and expandable. New articles can be added manually or via user submissions through the application interface, enabling continuous dataset growth and periodic retraining of the model. The design allows the system to adapt to evolving misinformation trends in the Bosnian digital media space. This modular design enables reproducible model updates and allows adaptation to evolving patterns of misinformation in the Bosnian digital media space.

VII. SURVEY METHODOLOGY AND RESULTS

An online survey was conducted among diverse demographic groups in Bosnia and Herzegovina to analyze news consumption habits, awareness of fake news, and attitudes toward AI-driven detection tools. The survey aimed to complement the technical analysis by providing insights into how citizens interact with digital news sources and perceive automated verification systems. The survey results indicate that a significant portion of respondents experience difficulties in reliably identifying fake news content. This finding supports the relevance of the proposed machine learning-based approach, as it highlights the need for automated tools that can assist users in detecting deceptive information.

[8] Jones, K. S. (1972). A statistical interpretation of term specificity

[9] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in Proc. Int. Joint Conf. on Artificial Intelligence (IJCAI), vol. 14, no. 2, pp. 1137–1145, 1995.

A. *Demographics and Consumption*

The majority of responses were individuals aged 18-34, identified as the primary consumers of online news. Gender distribution showed 60% female and 40% male participation. Most respondents hold a university degree, indicating a relatively educated sample. Findings show that news is consumed multiple times per day, primarily through social networks such as Instagram, Facebook and Tiktok, followed by visits to online news portal. These results suggest that younger populations are highly exposed to digital information streams where misinformation can easily propagate.

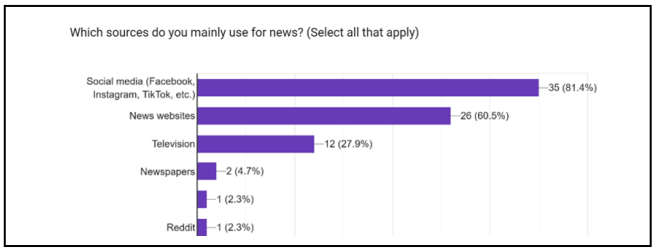

Figure 1. Demographic and news consumption profile of respondents

B. *Awareness and Attitudes*

Younger and more educated expressed high confidence in identifying fake news, while older participants remained more neutral or uncertain. Despite this, a significant portion of respondents admitted to sharing information later found to be false, particularly on sensitive topics such as politics or health. These finding highlight the discrepancy between perceived ability to identify fake news and actual behavior, underscoring the need for user-friendly verification tools that can be integrated into daily habits.

Figure 2. illustrates the responses to the question: "*Have you ever shared news online that later turned out to be false?"* The chart indicates that 60.5% of respondents never shared false news, 14% admitted occasional sharing, 4% shared frequently, and 20.9% were unsure. The distribution emphasizes that even among users confident in identifying fake news, accidental sharing still occurs, reinforcing the necessity for real-time AI verification systems and media literacy interventions.

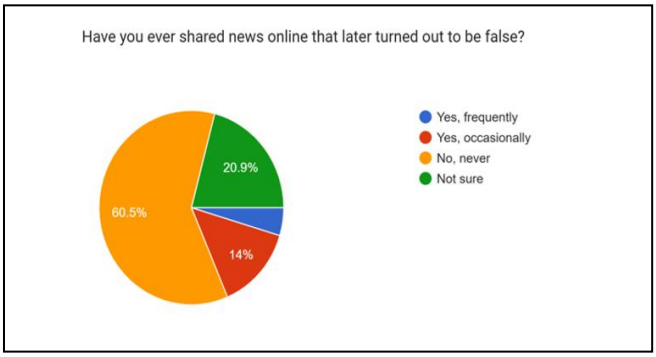

Figure 2. Awareness and Attitudes Toward Fake News

C. *Trust in AI and Desired Features*

Approximately 50% of respondents awareness of AI tools for news verification. While absolute trust remains low, there is strong willingness to utilize such tools prior to sharing content online. This indicates a potential for AI-assisted verification systems to influence user behavior positively, even among those initially skeptical.

Figure 3. presents the distribution of trust levels in AI tools for identifying fake news. The responses show that 34.9% were neutral**,** 32.6% somewhat trust**,** 18.6% do not trust much**,** 11.6% do not trust at all, and a small percentage expressed complete trust**.** This diversity highlights the necessity for AI systems to be not only accurate but also transparent and explainable to gain user confidence.

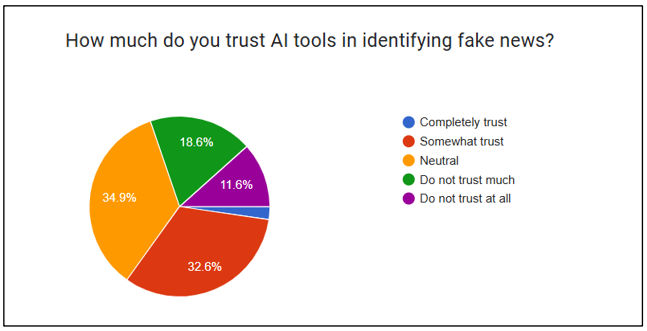

Figure 3.. Awareness and Trust in AI Tools

The results of survey are directly linked to the design of the proposed system described in Section VI. They inform both the technical development and the user interface, ensuring that the AI tools align with user expectations. For instance integrating probability scores, visual explanations, and alerts about suspicious content can address the concerns expressed by respondents, thereby bridging the gap between technical capability and user trust.

Furthermore, the findings underline the importance of educational interventions alongside AI deployment. Users' limited trust suggests that technological solutions alone are insufficient; they must be complemented by awareness programs to enhance understanding of how AI models

evaluate news credibility. This combination of technical rigor and user-centered design can maximize the system's effectiveness in reducing the spread of misinformation.

The results indicate that the proposed model achieves a balanced trade-off between precision and recall, as reflected by the F1-score. This balance is particularly important in the context of fake news detection, where both false positives and false negatives can have significant consequences. The achieved performance suggests that the selected feature representation and classification approach are suitable for identifying deceptive content in a low-resource language setting.

VIII. CONCLUSION

This research explores the problem of fake news in Bosnia and Herzegovina and provides a technical solution using AI and NLP. The study emphasizes the importance of local context and language, which are often neglected in global solutions. By developing a specialized dataset and a transparent classification model, this work strengthens media literacy and digital security. The results confirm that traditional methods, when adapted to regional linguistic specifics, provide an excellent foundation for future advancements in the domestic IT industry and media verification.

IX. REFERENCES

[1] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake news detection on social media: A data mining perspective," *ACM SIGKDD Explorations Newsletter*, vol. 19, no. 1, pp. 22–36, 2017.

[2] Raskrinkavanje.ba, "Fact-checking archive and methodology," [Online]. Available: https://raskrinkavanje.ba.

[3] S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," *Science*, vol. 359, no. 6380, pp. 1146–1151, 2018.

[4] E. C. Tandoc Jr, Z. W. Lim, and R. Ling, "Defining 'fake news': A typology of scholarly definitions," *Digital Journalism*, vol. 6, no. 2, pp. 137–153, 2018.

[5] H. Allcott and M. Gentzkow, "Social media and fake news in the 2016 election," *Journal of Economic Perspectives*, vol. 31, no. 2, pp. 211–236, 2017.

[6] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.

[7] H. Ahmed, I. Traore, and S. Saad, "Detection of Online Fake News Using N-Gram Analysis and Machine Learning," in *Lecture Notes in Computer Science*, vol. 10618, pp. 127–138, 2017.

[8] K. S. Jones, "A statistical interpretation of term specificity and its application in retrieval," *Journal of Documentation*, vol. 28, no. 1, pp. 11–21, 1972.

[9] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proc. Int. Joint Conf. on Artificial Intelligence (IJCAI)*, vol. 14, no. 2, pp. 1137–1145, 1995.

BIOGRAPHY

Lamija Bašalić is an ambitious tech enthusiast with a deep love for science and innovation, currently in her final, fifth year of master's studies in Software Engineering at Burch University. She built a strong foundation in computing after earning her undergraduate degree in Computer and Information Technologies at the University of Sarajevo, Faculty of Traffic and Communications. Her academic journey began at the Economic School in Vogošća, where she developed an early interest in systematic analysis. Lamija's research focuses on applying advanced machine learning techniques and natural language processing (NLP) to solve complex, data-driven challenges, including work on datasets for disinformation detection. She is dedicated to using technology to develop meaningful and useful real-world solutions.

Adnan Dželihodžić is a Professor and esteemed IT expert with a comprehensive background in both higher education and the technology industry. He currently holds faculty positions at the International Burch University, the Faculty of Traffic and Communications (University of Sarajevo), and the University of Zenica, Bosnia and Herzegovina.

Alongside his academic career, he is a key professional at Symphony, where he contributes to high-level software engineering projects and global technology solutions. His professional trajectory includes significant roles at the Central Bank of Bosnia and Herzegovina, where he gained deep expertise in mission-critical information systems Professor Dželihodžić is a prominent figure in the regional technical community, serving as a long-term jury member for the FIT Coding Challenge. He is widely recognized for his dedication to mentorship and his ability to bridge the gap between complex academic theory and industrial practice. His work focuses on advancing IT education and fostering the next generation of engineers, making him a pivotal link between the Bosnian academic sphere and the global IT market.